200601 - CEO - Software for Statistics and Optimization

**Coordinating unit:** 200 - FME - School of Mathematics and Statistics

**Teaching unit:**
- 1004 - UB - (ENG) Universitat de Barcelona
- 715 - EIO - Department of Statistics and Operations Research

**Academic year:** 2017

**Degree:** MASTER'S DEGREE IN STATISTICS AND OPERATIONS RESEARCH (Syllabus 2013). (Teaching unit Compulsory)

**ECTS credits:** 5

**Teaching languages:** Spanish

**Teaching staff**

**Coordinator:** KLAUS GERHARD LANGOHR

**Others:**
- Primer quadrimestre:
  - KLAUS GERHARD LANGOHR - A, B
  - ANTONIO MONLEON GETINO - A, B
  - LUIS ORTIZ GRACIA - A, B
  - ANA MARIA PÉREZ MARÍN - A, B

**Opening hours**

**Timetable:** At agreed times.

**Prior skills**

Concerning the R lectures, there will be two courses: an introductory-level course and an intermediate/advanced-level course. The first is for students with no or little experience of R, the second for students who have worked with R previously such as students with a degree in statistics. By contrast, the SAS lectures will be the same for all students.

**Requirements**

The intermediate/advanced-level R course requires that students have experience in working with R.

**Degree competences to which the subject contributes**

**Specific:**

3. CE-1. Ability to design and manage the collection of information and coding, handling, storing and processing it.

4. CE-5. Ability to formulate and solve real problems of decision-making in different application areas being able to choose the statistical method and the optimization algorithm more suitable in every occasion.

**Transversal:**

1. EFFECTIVE USE OF INFORMATION RESOURCES: Managing the acquisition, structuring, analysis and display of data and information in the chosen area of specialisation and critically assessing the results obtained.

2. TEAMWORK: Being able to work in an interdisciplinary team, whether as a member or as a leader, with the aim of
In this course, two statistical software packages are presented, R and SAS, that are widely used in the academic field as well as in business and industry.

The course aims to enable the student to use both software packages to
- read data from external files,
- carry out descriptive analysis,
- make high quality graphs to represent data,
- fit regression models to data sets,
- write own functions.

**Teaching methodology**

The lectures will take place in the computer room where both statistical packages, R and SAS, will be presented. The first part of the course will be dedicated to R and the second part to SAS. To illustrate the use of functions for statistics and graphics, real data sets will be used. During the course, students will have to do exams (in class) and a final exercise (at home) with each software package.

**Learning objectives of the subject**

In this course, two statistical software packages are presented, R and SAS, that are widely used in the academic field as well as in business and industry.

**Study load**

<table>
<thead>
<tr>
<th>Total learning time: 125h</th>
<th>Hours large group:</th>
<th>30h</th>
<th>24.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hours medium group:</td>
<td>0h</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>Hours small group:</td>
<td>15h</td>
<td>12.00%</td>
</tr>
<tr>
<td></td>
<td>Guided activities:</td>
<td>0h</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>Self study:</td>
<td>80h</td>
<td>64.00%</td>
</tr>
</tbody>
</table>
### Content

| **Introduction to R [Introductory level]** | **Learning time:** 1h 30m  
| | Theory classes: 1h  
| | Laboratory classes: 0h 30m |
| **Description:** |  
| a) The web page of R  
| b) Installation of R and its contributed packages  
| c) Sources of help |

| **R objects** | **Learning time:** 6h  
| | Theory classes: 4h  
| | Laboratory classes: 2h |
| **Description:** |  
| Creation and manipulation of  
| a) Numeric and alphanumeric vectors,  
| b) Matrices,  
| c) Lists,  
| d) Data frames |

| **Descriptive and exploratory analysis with R** | **Learning time:** 6h  
| | Theory classes: 4h  
| | Laboratory classes: 2h |
| **Description:** |  
| a) Reading external data files  
| b) Univariate descriptive analysis  
| c) Bivariate descriptive analysis  
| d) Graphical tools: histogram, box plot, scatter plot and others |
### Basic programming with R

**Learning time:** 6h  
Theory classes: 4h  
Laboratory classes: 2h

**Description:**
- a) Basic programming: loops with `for`, `while`, `if-else`
- b) Functions `tapply`, `sapply`, `lapply`
- c) Writing your own function
- d) Working with date variables

### Statistical inference with R: hypothesis tests and regression models

**Learning time:** 1h 30m  
Theory classes: 1h  
Laboratory classes: 0h 30m

**Description:**
- a) Hypothesis tests for one population
- b) Hypothesis tests for two or more populations
- c) Nonparametric tests
- d) Fit of general linear models

### Intermediate-level R topics

**Learning time:** 1h 30m  
Theory classes: 1h  
Laboratory classes: 0h 30m

**Description:**
- a) Reshaping data sets
- b) The Tidyverse packages
- c) Integrating R code in LaTeX documents
| **Introduction to SAS** | **Learning time:** 1h 30m  
Theory classes: 1h  
Laboratory classes: 0h 30m |
|-------------------------|----------------------------------|
| **Description:**        | a) Structure of the SAS programs: DATA and PROC.  
b) SAS data sets and libraries.  
c) Importation and exportation of data.  
d) Creation of variables. Commands of assignment.  
e) Merging databases.  
f) Management of data sets |

| **Basic procedures with SAS** | **Learning time:** 6h  
Theory classes: 4h  
Laboratory classes: 2h |
|-----------------------------|----------------------------------|
| **Description:**            | a) Introduction to procedures.  
b) Statistical and graphical procedures. |

| **Transformation and manipulation of data** | **Learning time:** 6h  
Theory classes: 4h  
Laboratory classes: 2h |
|---------------------------------------------|----------------------------------|
| **Description:**                            | a) Use of predefined functions.  
b) Conditional transformation of variables.  
c) Data generation with DO loops.  
d) Date variables.  
e) String functions.  
f) Error diagnosis and depuration. |
Qualification system

The final grade will be the average of the grades obtained in the different tests
a) with R (50%),
b) with SAS (50%).
Concerning R, there will be two exams in class (weight of each test: 30%) and a final practical work at home (weight: 40%). Concerning SAS, there will be two exams in class (weight of each test: 40%) and a final practical work at home (weight: 20%).
Bibliography

Basic:


Complementary:


